

波,用对照方法(第一遍参考信号,第二遍加二次谐波,第三遍加1/2次分谐波)对比试听,结果如表2。

表2 二次谐波与1/2次分谐波听感不可分辨阈值比较

频率 (Hz)	不可分辨阈值 (dB)	
	二次谐波	1/2 次谐波
4800	-20	-40
2000	-20	-40
1000	-20	-40

(3) 四分频、八分频的听感

对4000Hz、2000Hz、1000Hz 的正弦信号只加 3/4 次谐波,此时感觉到有两个音,但是不太好确定音高。

对 4000Hz、2000Hz、1000Hz 的正弦信号分别加上四分频(含 1/4 次、1/2 次、3/4 次)和八分频(含 1/8 次、1/4 次、3/8 次、1/2、5/8 次、4/3 次、7/8 次分谐波)。这时音杂粗糙,可听见

有好几个音混在一起。

五、结 论

扬声器非线性振动产生出分谐波已一再被证实,从所做的几组正弦信号分谐波听感实验可以看出,分谐波对音高、音质的影响是很明显的,对扬声器而言是增加了失真,不能高质量地重放体现原信号的本来面目。我们的实验证实了分谐波的幅度比高次谐波的幅度低得多时人耳就能察觉并且影响到音质,因此在设计、制作扬声器时是需要避免的。

本文在研究过程中得到了南京大学声学研究所陶擎天、倪皖菘、缪国庆老师的热枕帮助,在此表示衷心的感谢。

参 考 文 献

- [1] 魏表爵,陶擎天,电声技术 4(1986),1-4.
- [2] H. F. Olson, J. Acous. Soc. Am., 16-1 (1944), 1-4.

21-27

6

普通话韵母 /ao/ 与 /ou/ 的时频

协变对比分析

曹 剑 芬

(中国社会科学院语言研究所,北京-100732)

1992年12月2日收到

TN 912.3

汉语, 普通话, 韵母 ao, 韵母 ou, 语音信号处理

本文以语言所普通话语音数据库中十五个人的零声母音节声学参数为基础,对 /ao/ 与 /ou/ 这样一对极易相混的二合元音韵母内部频率域和时间域特性的协同变化进行微观分析。重点解析它们的频率特性随时间而协同变化的行为及其相关结构方面的异同,并给出相应的统计参数和相关的时频协变模式。

一、引 言

普通话里的二合元音 /ao/ 与 /ou/ 这对韵母具有许多相似之处。从感知特征来看,它们处于同一个相似度层级上^[1];就发音生理而言,尽管它们的音位目标值通常分别被标作 [au] 和

[ou];-但若从相关的声学元音动程图所反映的情况来看^[2],两者发音时舌位的动态变化的范围也比较接近;大致说来, /ao/ 是从后低元音 [a] 向着后半高元音 [o] 的方向滑移, /ou/ 是从后半低元音 [ɔ] 向着后高元音 [u] 的方向滑移,两者的舌位动程大致相当。因此,这是一对

极易相混的韵母，在感知听辨时具有一定的困难。我们曾经做过一个听辨试验，测试材料分为两个部分，第一部分共 40 个样品，其中包括五个发音人所发的自然音节 /ao/ 和 /ou/ 以及切除部分“头”或“尾”的 /ao/ 和 /ou/；这些样品分别依次排列，因此，相邻样品之间缺乏明显的对比；第二部分共 60 个样品，包括自然音节 /ao/ 和

/ou/ 以及从中切出的“头、腹、尾”各个长短不等的音段，并打乱次序随机排列，因此，这部分材料的相邻样品之间的对比是相当明显的。然后，请九个听音人分别听辨这两部分材料。第一部分是强迫听辨，要求听辨者必须判定哪个是 /ao/ 或更像 /ao/，哪个是 /ou/ 或比较像 /ou/；在第二部分的测试中，则要求听辨者听到甚么音

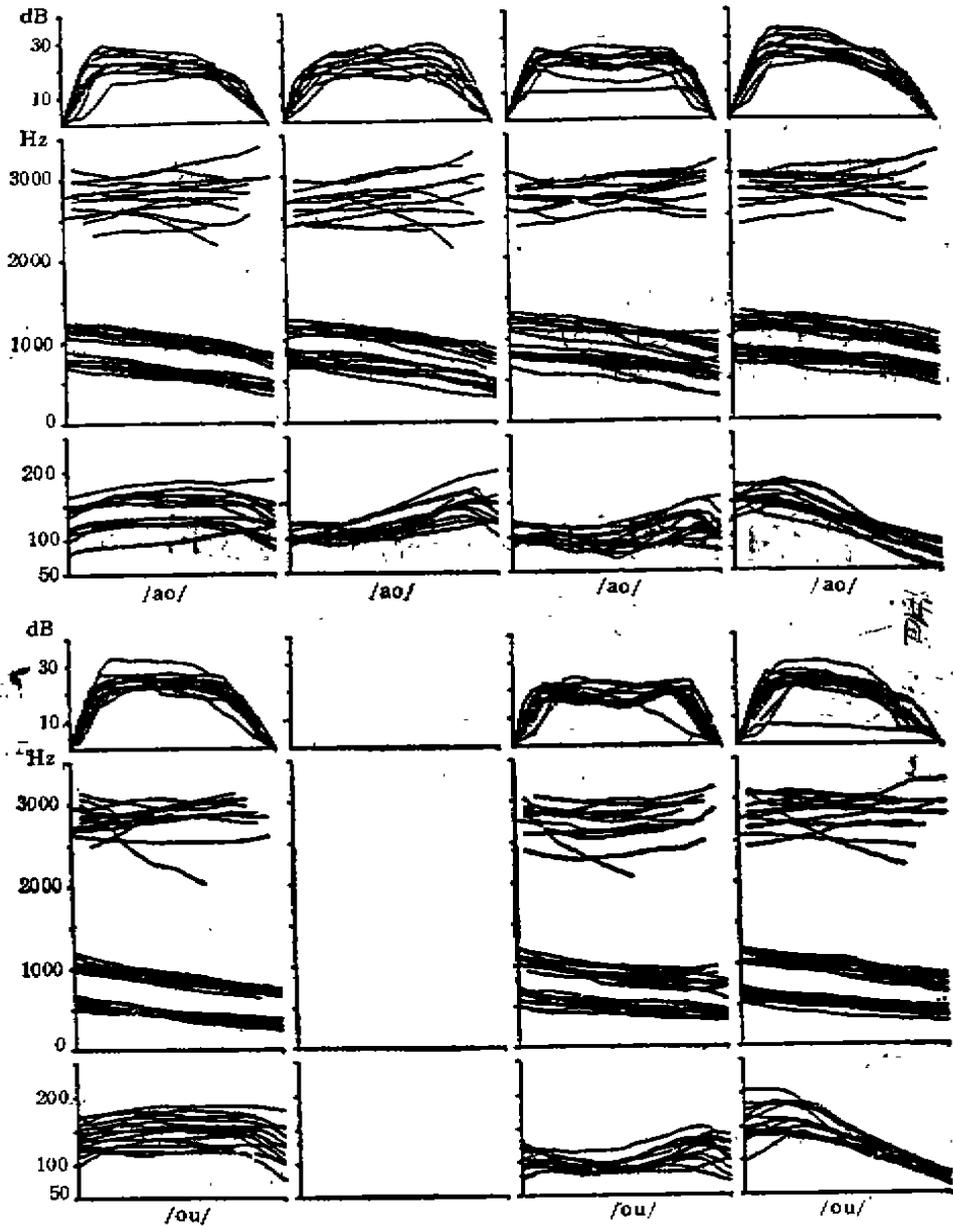


图1 普通话二合元音/ao/和/ou/零声母音节十五个人;时间归一的振幅、共振峰及音高模式图

就记甚么音。实验结果表明,在第一种条件、即相邻样品间没有明显对比的情况下, /ao/ 与 /ou/ 之间九个人的平均误判率竟高达 61.6%, 个别听音人的这项误判率甚至于高达 90%; 而在第二分部的测试中, 由于测试样品之间的对比非常明显, 这一项误判率就大大降低了, 但是, 仍有百分之 3% 左右的彼此误判, 上述情况说明, 在实际语言里, 人们可能多半是依靠对比性的语音环境来辨识这两个复合元音音节的, 而一旦孤立发音时, 即使由人耳来听辨, 也是极易被误判的。显然, 这种情况对于语音的计算机自动识别、尤其是孤立单词识别来说是个潜在的问题。不过, 作为普通话里两个彼此独立

的韵母, /ao/ 与 /ou/ 之间必定还存在着某种根本的区别特性。否则, 就不成其为两个独立的韵母了。

本文试图从语音声学信息分布的角度, 通过对这两个韵母内部频率域和时间域特性协同变化的行为的对比分析, 来考察它们彼此相似而又互相区别的本质特征。本研究以语言所的普通话语音声学参数数据库中零声母音节的材料为基础, 对这两个韵母十五个人的声学参数进行统计分析, 求出特征参量的统计平均值, 并给出相应的结构模式。研究的重点是探讨它们音色变化方面的定时关系(timing), 即从频域信息变化的时域分布的角度来考察它们的异同。

表 1 普通话 /ao/ 和 /ou/ 的一、二、三共振峰频率 (Hz) 的统计平均值和标准偏差

音节 测量点 (%)	/ao/									/ou/								
	F1			F2			F3			F1			F2			F3		
	mn	sd	n	mn	sd	n	mn	sd	n	mn	sd	n	mn	sd	n	mn	sd	n
0	778	70	60	1189	78	35	2671	196	9	615	73	45	1096	68	29	2729	116	8
5	769	67	60	1183	68	39	2746	181	44	590	68	45	1073	66	45	2819	182	39
10	760	66	60	1174	67	60	2736	183	55	576	65	45	1049	63	45	2803	173	42
15	751	66	60	1165	68	60	2736	185	59	558	63	45	1027	62	45	2796	179	44
20	741	65	60	1155	69	60	2737	182	59	545	60	45	1006	62	45	2787	172	44
25	733	66	60	1144	70	60	2740	178	60	532	57	45	986	63	45	2780	178	44
30	724	68	60	1130	70	60	2744	178	60	520	56	45	967	64	45	2773	186	44
35	715	70	60	1116	69	60	2749	178	60	506	55	45	949	66	45	2776	194	45
40	704	73	60	1102	68	60	2754	179	60	493	55	45	932	70	45	2776	203	45
45	691	75	60	1087	67	60	2760	181	60	479	56	45	915	71	45	2776	209	45
50	678	77	60	1071	67	60	2766	184	60	464	56	45	896	73	45	2771	217	43
55	664	79	60	1053	66	60	2774	188	58	449	54	45	878	73	45	2774	226	43
60	651	82	60	1034	68	60	2776	193	58	435	54	45	860	71	45	2796	208	42
65	636	85	60	1014	70	60	2776	204	55	424	55	45	843	70	45	2792	218	40
70	618	87	60	993	71	60	2776	212	55	415	55	45	825	68	45	2821	189	37
75	596	87	60	970	72	60	2791	215	53	406	55	45	809	67	45	2822	197	36
80	575	85	60	944	74	60	2806	218	50	397	55	45	794	67	45	2828	211	34
85	554	86	60	918	77	60	2822	205	46	387	56	45	780	69	45	2818	193	24
90	532	86	60	890	81	60	2849	216	33	377	56	45	770	73	44	2846	177	18
95	506	87	60	861	85	58	2875	239	21	369	56	45	753	72	39	2920	196	8
100	480	90	60	833	88	43	3148	133	3	364	54	45	728	66	26	2997	208	3

音节平均时长 (ms): /ao/ = 336

/ou/ = 326

注: mn: 平均值, sd: 标准偏差, n: 样本数

二、韵母/ao/与/ou/的相似性:

从声学特性来看, /ao/与/ou/的声谱外观非常相似。图1是根据十五个人的、时间归一的基频、共振峰和振幅的测量值所作的/ao/和/ou/(包括四声)的声谱模式图。

从这个图上我们可以大致看出, /oa/与/ou/在音色方面的相似性具体表现在两个方面: 第一, 它们的第一、二共振峰(F1, F2)都是持续地平行下滑, 第三共振峰(F3)除个别发音人以外, 总的趋势都是相对平稳, 而尾部微升; 第二, 它们的共振峰结构, 即各个共振峰之间的相对关系也大致相当。

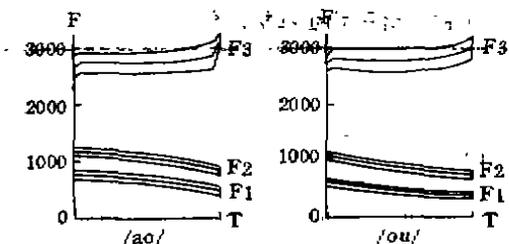


图2 普通话十五个发音人的 /ao/ 和 /ou/ 时间归一的共振峰平均模式图

图2是上述十五个人的/ao/和/ou/的时间归一的共振峰平均值及其标准偏差的模式图, 它更加清楚地表明了这两个二合元音在声学上的近似性: 首先, 它们的F1和F2都是分别从1000Hz上下平行下滑, F3都是在2500到3000Hz这个范围内变化; 其次, 无论是/ao/还是/ou/, F1与F2或F2与F3之间的频率差值也都不相上下。例如, 根据我们的统计结果(见表1)来看, 它们的F1与F2之间的平均频率差值几乎都在400Hz上下。

三、/ao/与/ou/的区别性

1. 信息分布的频率域区别特性:

/ao/与/ou/的共振峰起始频率值以及相应的共振峰之间的频率差虽然看起来很相近, 而

实际上还是有区别的。尽管这种区别不如与其它韵母之间(例如 /ao/ 与 /ai/, /ou/ 与 /wo/, /ao/ 与 /ya/, 等等)的区别来得明显, 而且, 还存在着明显的个体差异。同时, 就共振峰的绝对值而言, 不同个体的 /ao/ 与 /ou/ 之间甚至还可能出现某些彼此交叠的现象。但是, 从总体上看, /ao/与/ou/ 的共振峰频率值是各不相同的。这种区别性可以从表1所提供的相关数据中大致窥得。表1的数据是根据十五人的原始声谱图测得的共振峰数据、经过时间归一处理而获得的21个测量点上的统计平均值和标准偏差。相关的变量分析表明, 它们之间的这种差异的显著性水平很高 ($P < 0.001$)。尽管对于感知听辨来说, 这种差异也许并不是唯一的影响因素, 但毕竟是个不可忽视的重要方面。例如, 在上述的感知试验中, 我们发现, /ao/与/ou/之间的误判, 几乎都是从/ao/中切出的后半部分被误判为/ou/, 这部分误判率高达60%强, 而从/ou/中切出的部分被误判为/ao/的只有两例, 仅占0.8%。这种偏颇很可能就是由于/ao/的后半部分各个共振峰的频率值及其相对频率差正好跟/ou/的比较相近的缘故。

2. 信息分布的时间域区别特性:

越来越多的相关研究发现^[3-4], 复合元音音色的变化实际上反映了在时间轴上有定的、即有计划的(planned)发音动作的变化。同样, 汉语里也有许多证据说明, 语音特性的时域结构特点是决定它们音质(timbre)特性的一个至关重要的因素。例如, 普通话里的 /er/ 通常被列入单元音之列, 但是, 若从声学表现来看, 它的共振峰不但具有动态变化, 而且, 其共振峰频率随时间而协同变化的行为还因声调而不同, 而且, 合成试验也表明, 这对于合成音质很要紧。由此可见, 对于复合元音来说, 它们的时域结构特点就更是不可忽视的关键了。在这方面, 我们关于二合元音 /ai/ 的汉英对比分析已经为此提供了一个相当生动的例子。由此不难推测, /ao/与/ou/ 这两个韵母之间的语音边界很可能主要同它们内部信息分布的时域结构特性有关。事实上, 如若仔细观察图2的共振

峰平均模式，便会发现两者共振峰滑移的轨迹的确存在着微妙的差异：*/ao/* 的 F1 和 F2 平行下滑的轨迹类似于“下弦月”形，而 */ou/* 的则略呈“上弦月”形。这种滑移轨迹的细微差异看起来并不起眼，但是，进一步的微观分析发现，这种微妙的外观差异却正反映了它们在频时协变关系方面、即共振峰频率变化的时间结构方面不可忽视的重要区别。

为了剖析和比较这两个二合元音动态特性

的微观方面，我们根据表 1 所提供的 F1 和 F2 的统计平均值和音节的平均时长，首先分别计算了 */ao/* 与 */ou/* 第一和第二共振峰滑移变化的总体斜率，然后依次计算每一个测量点相对于前一个测量点之间的共振峰滑变斜率及其与总体斜率之离差。考虑到起始和收尾两个测量点上实际样本数的不足，这里只统计了其余十九个测量点之间的参数，具体数据如表 2 所示。

表 2. */ao/* 与 */ou/* 的 F1, F2 各相应测量点的滑变斜率及其与总体斜率之离差

测量点	<i>/ao/</i>				<i>/ou/</i>			
	F1		F2		F1		F2	
	斜率	离差	斜率	离差	斜率	离差	斜率	离差
1	0.54	-0.28	0.18	-0.85	1.23	0.44	1.41	0.32
2	0.54	-0.28	0.54	-0.49	1.17	0.38	1.47	0.38
3	0.54	-0.28	0.54	-0.49	1.10	0.31	1.35	0.26
4	0.60	-0.22	0.60	-0.43	0.80	0.01	1.29	0.20
5	0.48	-0.34	0.65	-0.38	0.80	0.01	1.23	0.14
6	0.54	-0.28	0.83	-0.20	0.74	0.05	1.17	0.08
7	0.54	-0.28	0.83	-0.20	0.86	0.07	1.10	0.01
8	0.65	-0.17	0.83	-0.20	0.80	0.01	1.04	-0.05
9	0.77	-0.05	0.89	-0.14	0.86	0.07	1.04	-0.05
10	0.77	-0.05	0.95	-0.08	0.92	0.13	1.17	0.08
11	0.83	0.01	1.07	0.04	0.92	0.13	1.10	0.01
12	0.77	-0.05	1.13	0.10	0.86	0.07	1.10	0.01
13	0.89	0.07	1.19	0.16	0.67	-0.12	1.04	-0.05
14	1.07	0.25	1.25	0.22	0.55	-0.24	1.10	0.01
15	1.31	0.49	1.37	0.34	0.55	-0.24	0.98	-0.11
16	1.25	0.43	1.55	0.52	0.55	-0.24	0.92	-0.17
17	1.25	0.43	1.55	0.52	0.61	-0.18	0.86	-0.23
18	1.31	0.49	1.67	0.64	0.61	-0.18	0.61	-0.48
总体斜率: 0.82			1.03		0.79		1.09	

从表 2 最末一行的数据所代表的 */ao/* 与 */ou/* 各自的 F1 和 F2 滑移的总体斜率来看，都是在 0.80 和 1.06 左右，两者并无明显差异，所以，粗看起来它们的声谱外观很相似；可是，若认真比较表 2 所给的各个测量点上的具体滑移斜率及其与总体斜率之离差，就可看出，它们

的共振峰滑移变化的时间分布行为显然是不同的，就 */ao/* 而言，无论是 F1 还是 F2，约在全程的 60% 以内，其滑变斜率都低于或接近于总体斜率，此后则明显地高于总体斜率；而 */ou/* 的情况几乎正好相反。图 3 正是形像地反映了两者在频时协变结构方面的这种重要区

别。由此我们可以直观地看出，两者根本的区别就在于它们频时协变的行为不一样：*/ao/*的共振峰随时间协同变化的速率是先低后高，而

*/ou/*的变化速率是先高后低。这种微妙而又不可忽视的区别特性从它们的声谱上通常是不易发现的。

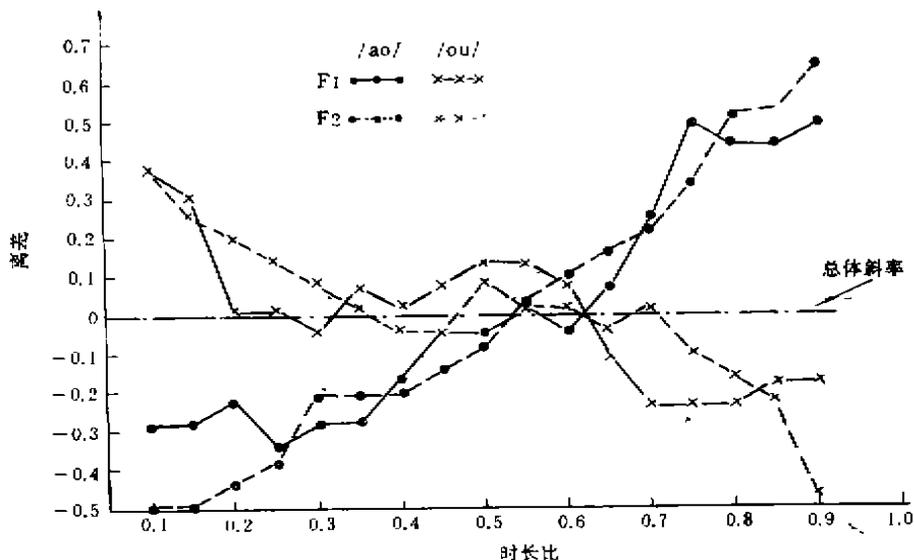


图3. 普通话*/ao/*与*/ou/*的时频协变结构比较

表3. */ao/*与*/ou/*的第一共振峰滑移斜率差异的 T-检验结果

测量点	t(1,2)	df	p
1	5.09	63.98	3.64D-05
2	-5.08	94.57	2.87D-05
3	2.37	69.49	1.09D-02
4	2.17	73.86	0.0312
5	1.98	96.92	4.77D-02
6	1.37	75.95	0.1713
7	-3.38	95.78	1.42D-03
8	-4.96	103.8	3.5D-05
9	-5.44	103.49	1.34D-05
10	-3.49	68.84	1.19D-04

表3的数据是一组对上述差异进行 T-检验的结果。从这些数据可以看出，绝大部分测量点上差异的风险率 (P) 都低于千分之一，说明两者在这方面的差异的显著性水平很高。不过，值得注意的是，处于音节中段的若干测量点上的差异不够显著，这就说明，这部分是这两个

二合元音的区分低谷。了解这一点也许不无意义。因为对于复合元音来说，其复合信息的主要负载体是它们的渐变中段^④，而*/ao/*与*/ou/*的区分低谷恰恰落在了这个关键的部分，这或许就是这两个元音在单独听辨时那么容易相混的一个根本原因。

四、小 结

/ao 与 /ou/ 的对比分析说明, 它们的根本区别在于各自的频时协变行为的不同。这个例子起码给我们提供了以下两点启示: 第一, 自然语音的动态特性不仅取决于频率变化的空间分布, 而且受频率变化的时间分布格局的制约。因此, 必须把这两方面结合起来, 对它们协同变化的微观结构加以考察, 才能全方位地了解它们的动态变化的规律及其相关的结构模式; 第二, 复合元音共振峰的动态变化既不是自始至终均匀的线性滑移, 也不是两个或三个目标值之间的随意滑动, 而是共振峰频率在时间轴上有定的、即遵循一定规律的滑移变化, 因此, 要真正认识复合元音的区别特性, 仅仅知道它们的共振峰起迄频率和大致的滑移变化方

向还是远远不够的。相对说来, 更为重要的是应该搞清楚它们各自特有的频时协变规律及其相关的结构模式。同时, 这种模式随语言或方言的不同而不同, 必须对具体语言作具体分析。所以, 真正了解和掌握这些特定的模式, 必将有助于合成语音音质的改善和自动识别准确率的提高。

参 考 文 献

- [1] Zhang Jialu, Lü Shinan, Qi Shiqian, *Journal of Chinese Linguistics*, 10-2 (1982), 189-206.
- [2] 吴宗济,《汉语普通话单音节语图册》, 中国社会科学出版社, 1986.
- [3] Hongmo Ren, on the acoustic structure of diphthongal Syllables, UCLA Working Papers 65 (1986).
- [4] Willicm. J. M. Peeters, Diphthong Dynamics, Dissertation, University of Utrecht, The Netherlands, 1991.
- [5] 曹剑芬、杨顺安, 北京话复合元音的实验研究,《中国语文》, 6(1984), 426-433.

27-34

汉语, 三合元音, 普通话, 声学, 最小时间感知阈

普通话三合元音音节最小时间感知阈及其声学特性

祖 清 清

(中国社会科学院语言研究所 北京 100732)

1992年12月31日收到

TN 912.3
11027

本研究的实验材料取自中国社会科学院语言研究所语音数据库。库中存有 15 个男音的语音材料, 共有 $15 \times 15 = 225$ 个三合元音音节。本研究的主要目的是从普通话三合元音入手, 在对 15 个说话人的语音材料统计的基础上, 通过对最小时间感知阈 T_{lim} 的测量与研究, 从声学 and 感知的角度, 给出三合元音必不可少的信息, 指出多余信息。

实验结果表明, T_{lim} 内的共振峰变化情况可分为两类。一是动态特性, 它的表现是: (a) $\Delta F_1 > 90\%$, ΔF_2 约 50%; (b) T_{lim} 内至少包括 F_1, F_3 两个拐点中的一个; (c) T_{lim} 内包括 F_2 变化最剧烈的部分。这四点对四个三合元音是一致的。第二类是边界条件, T_{lim} 受到位置和大小两方面的限制, 证明其边界共振峰频率十分重要。

一、引 言

过去对于复合元音的研究, 无论其出发点
应用声学

是从声学、生理、感知哪一个角度, 所得出的结论都是: 复合元音是有动程的元音, 需要用目标值和过渡两方面的参数描述。

早在六十年代, Lehiste 等^[1]、Holbrook 等^[2]